ALGORITHMS FOR SPECTRAL DECOMPOSITION WITH APPLICATIONS TO OPTICAL PLUME ANOMALY DETECTION

ASHOK N. SRIVASTAVA*, PH.D., BRYAN MATTHEWS, AND SANTANU DAS, PH.D.

Abstract. The analysis of spectral signals for features that represent physical phenomenon is ubiquitous in the science and engineering communities. There are two main approaches that can be taken to extract relevant features from these high-dimensional data streams. The first set of approaches relies on extracting features using a physics-based paradigm where the underlying physical mechanism that generates the spectra is used to infer the most important features in the data stream. We focus on a complementary methodology that uses a data-driven technique that is *informed* by the underlying physics but also has the ability to adapt to unmodeled system attributes and dynamics. We discuss the following four algorithms: Spectral Decomposition Algorithm (SDA), Non-Negative Matrix Factorization (NMF), Independent Component Analysis (ICA) and Principal Components Analysis (PCA) and compare their performance on a spectral emulator which we use to generate artificial data with known statistical properties. This spectral emulator mimics the real-world phenomena arising from the plume of the space shuttle main engine and can be used to validate the results that arise from various spectral decomposition algorithms and is very useful for situations where real-world systems have very low probabilities of fault or failure. Our results indicate that methods like SDA and NMF provide a straightforward way of incorporating prior physical knowledge while NMF with a tuning mechanism can give superior performance on some tests. We demonstrate these algorithms to detect potential system-health issues on data from a spectral emulator with tunable health parameters.

1. Introduction

The analysis of spectral signals is one of the classic problems in physics. Numerous references, dating back to the 17th century have discussed optical spectra, and then with the deeper understanding of quantum mechanics, the relationship between chemical elements and spectral energy. For the purposes of this paper, we model the observed spectral data as a time series of spectra $Y(\Lambda, N)$. The columns in this matrix correspond to the observations of the spectra at a given time N. The rows correspond to the wavelengths at which the spectral observations are made. The spectral components at a given time are a vector of

*NASA Ames Research Center, Moffett Field, CA, 94035, ashok.n.srivastava@nasa.gov.

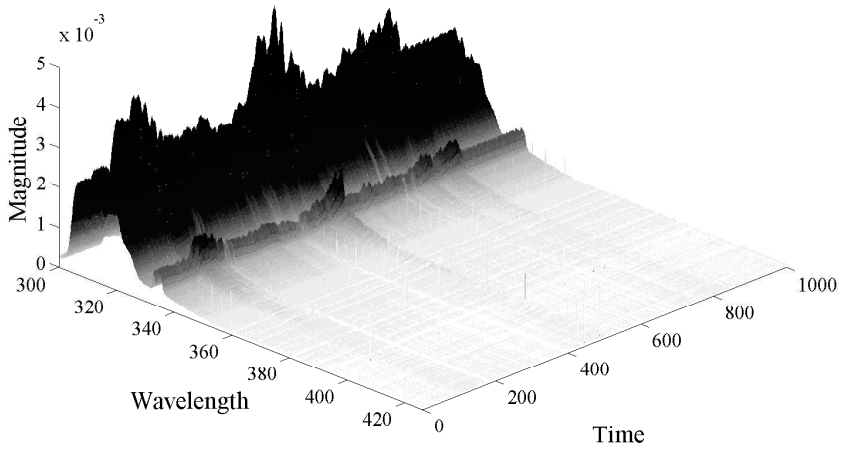


FIGURE 1. This 'waterfall' figure shows a typical spectral signal generated by our Spectral Emulator, which is a software program that generates spectral time series with known statistical properties. The output of the Spectral Emulator has some properties that are similar to emissions spectra from liquid propulsion systems. The figure shows significant structure in the lower wavelength bands and has been seeded with known wavelength locations for 10 elements. Notice that the spectral signatures are time varying in nature.

observations of length Λ where Λ depends on the resolution of the spectral data acquired by the detector. In our current application, Λ is typically 1061.

The problem that we address in this paper is to develop and test approaches to extract relevant system-health information from $Y(\Lambda, N)$. We advance this by studying various matrix factorization techniques which result in signals that are of lower dimension and that can contain relevant health information. The standard approach to solve this problem is to use Principal Components Analysis (PCA) which results in a factorization of the matrix into a set of m orthogonal basis vectors where $m \ll \Lambda$ and m is chosen from the eigenspectrum of Y. As we will see in our analysis, these results can be very useful in understanding the underlying data generating process. However, the PCA algorithm suffers from some shortcomings that require us to make further algorithmic advances. These deficiencies will be discussed later in this manuscript.

Figure 1 shows a few columns in a Y matrix that is generated from a Spectral Emulator that is discussed later in this paper. The key properties of spectral signals are that they exhibit variations in multiple wavelength bands (left-hand axis) that can be correlated due to the underlying data generating process. The large signature on the left-hand side of the figure is often indicative of the hydroxide burn process that covers a large band of wavelengths. Although from a statistical perspective Y can be modeled as a multivariate time series, it is helpful to note that the spectral properties are highly correlated with one another.

Most data-driven anomaly detection algorithms are not able to directly operate on high dimensional data sets because of the so-called curse of dimensionality [7]. The spectral decomposition methods we discuss

here result in significant dimensionality reduction while preserving a significant amount of systems health related information as measured by the performance of the detection algorithm. This paper shows, however, that standard dimensionality reduction techniques, such as PCA must be applied judiciously in situations where the amount of data is significant or when one has a priori knowledge.

2. Spectral Emulator

The emulator is designed to generate a set of time series of spectra similar to what is measured with a spectrometer in optical plume analysis in liquid propulsion engines. It does not employ a physics based model for data generation, but instead the intent is to emulate similar signals that can be found in optical plume data with the assumption that the basis vectors are mixed. Since the decomposition techniques we are dealing with operate under the assumption that the signals are combinations of basis vectors we are restricting the emulator to a set of spectral vectors S and the corresponding temporal profiles a, with a noise factor η. The combination is as follows: $Y = Sa + \eta$. The spectral basis vectors consist of three distinct components

1.) The estimated broadband spectral profile of a hydroxide burn.
2.) The spectral profile due to background scattering.
3.) Ten unique "severity one" elemental wavelengths profiles.

The hydroxide (OH) component represents the spectral features produced from the burning of pure hydrogen and oxygen during engine operation. It makes up the majority of the energy in the signal and has a broadband spectral profile. The emulator recreates this spectral feature by building a higher order polynomial function with coefficients that are allowed to vary from run to run in such a way as to have a similar patter as the OH burn found in the real data. The corresponding time profile for the OH component can either be generated as a linear slope or exponential profile with intermittent amplitude changes or any combination of the above. The background scatter profile attempts to represent the phenomena that occurs when particles produce radiation which is reflected amongst the rest of the particles in the plume. This creates a background noise that has a somewhat periodic characteristic in the spectral domain [1, 2]. The emulator recreates this spectral feature by using a weighted combination of sines and cosines to produce the desired effect. The corresponding background scatter

The elemental profiles each have a set of primary and secondary wavelengths that correspond to known severity one list elements found during engine operation (Ni, Fe, Cr, Co, Cu, Mn, Ca, Al, Ag, and Mg) [6]. The spectral profile is recreated by generating a high signal to noise ratio at the peaks of the primary and secondary wavelengths for a given element and assigning positive uniform noise to the remaining wavelengths in the spectrum. The corresponding time profiles for the elements have intermittent spikes and have a baseline close to zero. This behavior has been observed and documented in previous studies [1, 2, 5, 4]. In these studies spikes due to the intermittent presence of elements have been identified and confirmed and we have also been able to observe similar traces in the real data after applying the decomposition techniques mentioned in this paper.

$$Y_{Final} = \sqrt{\sum_{i=1}^{m} Y_i^2 \cdot \omega_i} + \eta. \tag{1}$$

Y_i represents the time series of spectra matrix for each component including the OH, the background scatter, and each of the ten element components. Here ω_i corresponds to the energy weight for the i^{th} component. The element and background scatter components are each given very small weights, leaving the remaining energy for the OH component.

TABLE 1. The table shows the prominent spectra lines for SSME elements in the spectral range of 320 to 426 nm [17]. The wavelength with an asterick denotes the dominant wavelength among all the spectral lines corresponding to that element.

Elements	Wave lengths
Nickel	341.5, 345.9, 346.2, 349.3, 351.6, 352.6*, 362.0
Iron	372.0, 373.7, 374.6, 375.0, 382.1, 382.6, 385.7, 386.1*, 388.7, 388.7
Chromium	357.9, 359.4, 360.6, 425.6*
Cobalt	341.3, 345.0, 345.5, 346.6, 347.5, 350.3, 351.5, 353.1, 357.6, 387.4*
Copper	324.8*, 327.4
Manganese	403.4*
Calcium	422.6*
Aluminium	396.1*
Silver	328.0, 338.3*
Magnesium	370.2*, 371.9, 380.8, 383.3, 384.5

After combining all contributions and their appropriate energy weighting from the individual Y matrices the energy of Y_{Final} has unit energy. The resulting Y_{Final} matrix has dimensions $\Lambda \times N$ where Λ is the number of wavelengths in the spectral domain and N is the number of time samples. At any given time sample the spectra contains a mixing of all components. The decomposition techniques addressed in this paper attempt to extract these basis vectors and isolate the element components. Unlike test stand data the emulated data contains known ground truth for all samples in time that correspond to the element burns and therefore we can compute detection rates, which are reported in the results section.

3. Decomposition Algorithms

The main idea behind the decomposition algorithms discussed here is to reduce the number of dimensions in the observed signal to extract features that can be used for anomaly detection. Ideally the features would be indicative of the health of the system under study. For our examples, we assume that we are studying data from a liquid propulsion system such as the space shuttle main engine. These extracted signals should correspond to known chemical species in the propulsion system. Higher concentrations of certain metals, such as Cr, Ni, and Fe can be indicative of adverse conditions in the engine [17]. Thus, these algorithms must generate interpretable signal decompositions so that users can have a clear understanding of the underlying physical mechanism. This 'interpretability' requirement is not necessarily achieved by standard statistical algorithms. This paper overviews some key innovations in the statistical machine learning community that can be useful for this application domain [1].

SPECTRAL DECOMPOSITION ALGORITHM

The approach that we take to decompose the spectral time series $Y(\Lambda, N)$ utilizes methods in the blind source separation literature [11]. In so-called blind source separation problems, we assume that a set of stationary signals S is mixed through a linear mixing matrix a. The result of this mixing matrix is the observed signal Y [16]:

[1]More detailed information on these algorithms and OPAD application is available at Dashlink website (http://dashlink.arc.nasa.gov/).

$$\mathbf{y}(t) = \sum_{i=1}^{m} S_i a_t + \eta_t \quad (2)$$

$$Y = \mathbf{Sa} + \eta \quad (3)$$

In this formulation, $\mathbf{y}(t)$ is a column vector of size $\Lambda \times 1$, S_i is a vector of $\Lambda \times 1$ and a_t is a vector of size $1 \times m$ and we assume that there are a total of m stationary components of S. We will need to solve for $\mathbf{S}$ and $\mathbf{a}$ given the Y vectors. The procedure to do this decomposition is given in [16] and is called the Spectral Decomposition Algorithm (SDA). SDA works by assuming a random starting point for S, computing a and then re-computing a given the current estimate of S. The update equations are based on a least squares solution to the problem and provides fast convergence to a solution and are given in the cited reference. The cost function that is minimized in SDA is $||Y - \mathbf{Sa}||^2$ which is solved using an unconstrained optimization procedure.

This algorithm features an easy way to incorporate prior knowledge. For example, suppose that one knows the spectral emission lines for e_{il} elements under study as found in Table 1, where 'i' corresponds to a particular element and 'l' correspond to the specific wavelengths for a spectral profile. In this case, these e_{il} spectral lines can be encoded into the initial guess of the matrix S. The first estimate of a then will result in the optimal (in the least squares sense) result given the initial guess. Subsequent iterations will lead to further refinements in the initial estimates of S with additional signals being estimated, i.e. if $m > max(i)$. With a random initialization, this algorithm converges to a set of orthogonal stationary signals. While there are many solutions to the linear model show in Equation 3, SDA is particularly fast and flexible in its formulation, thus providing an ideal model for decomposing spectra.

A key weakness of the algorithm, however, is that it assumes that the mixing between S and a is linear. In many real-world cases, it may be the case that a nonlinear mixing occurs. Depending on the nature of the nonlinearity, SDA may not correctly capture the appropriate components. A nonlinear form for this mixing can be expressed as $Y = \Phi(S, a) + \eta$. We have solved the problem for Φ being a linear operator.

PRINCIPAL COMPONENTS AND FACTOR ANALYSIS

The solution can be shown to be a variant of the famous PCA algorithm [10] developed originally by Hotelling in the 1930's. The principal components algorithm computes an orthogonal decomposition of the correlation matrix produced by the matrix Y. In this computation we define the correlation matrix $\Sigma = (Y - \mathbf{j}\bar{Y})^T(Y - \mathbf{j}\bar{Y})$, where $\bar{Y}$ is the mean of the data in the columns of Y and $\mathbf{j}$ is a vector of size $N \times 1$ where we assume that we have N spectral samples. A diagonalization of this matrix yields the principal components:

$$\Sigma = V^T \Delta V \quad (4)$$

where first m columns of the matrix V correspond to the largest eigenvalues in the diagonal matrix Δ. These eigenvectors can be easily shown to span the directions of maximum variance in the data matrix Y. This specific and unique property of PCA makes the stationary signals easy to interpret from a mathematical perspective. However, these components may not be easily interpretable from the point of view the data generating process. The PCA algorithm cannot be easily initialized with prior knowledge since the extracted signals are uniquely determined by Equation 4. PCA, as in the case of SDA, does not extract appropriate signals in cases with nonlinear mixing.

Input: $Y_{\Lambda \times N}$, m (desired rank),$\mathbf{S}_{\Lambda \times m}$, $\mathbf{a}_{m \times N}$ and Q (stopping criteria)
Step 1: Randomly initialize $\mathbf{S}$, $\mathbf{a}$ with positive values.
Step 2: While(not Q)
a) Update $\mathbf{a}$:=$\mathbf{a}.*(\mathbf{S}^T Y)./(\mathbf{S}^T \mathbf{S}\mathbf{a})$;
b) Update $\mathbf{S}$:=$\mathbf{S}.*(Y\mathbf{a}^T)./(\mathbf{S}\mathbf{a}\mathbf{a}^T)$;
end
Output: $\mathbf{S}$, $\mathbf{a}$

FIGURE 2. Steps of Standard NMF Algorithm.

In the statistical community, there is probabilistic a variant of PCA known as Factor Analysis, which is widely used in the social sciences. In Factor Analysis, the matrix Y is decomposed as $Y = \mathbf{S}\mathbf{a} + \delta_i$ where the factors S and a are assumed to have zero mean, orthogonal and of unit length (orthonormal). There are additional constraints placed on S and a in this decomposition. Once a decomposition is performed, it is possible to rotate the resulting factors S via a rotation matrix. This allows the analyst to identify features that may be more interpretable [10] [14].

NON-NEGATIVE MATRIX FACTORIZATION

In SDA and PCA, the matrix decompositions allow the elements of S and a to be either positive or negative. However, the spectral data that is observed in Y is always non-negative. Recently [9, 8, 12] a new matrix decomposition algorithm has been developed called non negative matrix factorization (NMF) which finds a decomposition $Y = \mathbf{S}\mathbf{a}$ such that all the elements of $\mathbf{S}$ and $\mathbf{a}$ are non negative. This decomposition preserves the important non-negative property of the spectral data and can lead to superior results in some cases.

NMF minimizes the squared reconstruction error $C = ||Y - \mathbf{S}\mathbf{a}||^2$ given the constraints that S and a contain non negative values. In some variants of the algorithm, it is possible to place a sparseness constraint on the solution matrices. This can lead to better and more interpretable solutions [9, 8]. Another attractive feature of NMF is that it converges rapidly and is easily interpretable for some applications. The pseudo code of some variants NMF algorithms using various updating rules can be obtained in the following review literature [15]. Like SDA, this constrained optimization problem leads to an iterative algorithm to update $\mathbf{S}$ and $\mathbf{a}$. The standard NMF algorithm is given in Figure 2.

NON-NEGATIVE MATRIX FACTORIZATION WITH ENERGY MINIMIZATION

We explored two novel variants of NMF that allows us to impose a further constraint on the energy (i.e., the sum of the squared values in the components of $\mathbf{S}$ and $\mathbf{a}$). These variants can be captured through the following two optimization functions:

$$C_1 = \frac{1}{2}||Y - \mathbf{S}\mathbf{a}||^2 + \frac{1}{2}\alpha_1||\mathbf{S}||^2 \tag{5}$$

$$C_2 = \frac{1}{2}||Y - \mathbf{S}\mathbf{a}||^2 + \frac{1}{2}\alpha_1||\mathbf{S}||^2 + \frac{1}{2}\alpha_2||\mathbf{a}||^2 \tag{6}$$

For cost function C_1, the objective is to minimize the squared reconstruction error (Euclidean distance) given the non-negativity constraints. The penalty function includes a function term that represents the energy of the hidden components in the spectral domain. Since the properties of the data are such that

Input: $Y_{\Lambda\times N}$, m (desired rank),$\mathbf{S}_{\Lambda\times m}$, $\mathbf{a}_{m\times N}$, Q (stopping criteria),α_1 and α_2 (regularization parameter)
Step 1: Intialize **S**,**a**.
Step 2: While(not Q)
a) Update $\mathbf{a}$:=$\dfrac{(\mathbf{S}^T Y)}{(\mathbf{S}^T\mathbf{S}+\alpha_2 I)}$;
b) $\mathbf{a} = \mathbf{a}. * (\mathbf{a} >= 0)$;
c) Update $\mathbf{S}$:=$\dfrac{(\mathbf{a}^T Y)}{(\mathbf{a}\mathbf{a}^T+\alpha_1 I)}$;
d) $\mathbf{S} = \mathbf{S}. * (\mathbf{S} >= 0)$;
end
Output: **S**, **a**

FIGURE 3. Pseudo code for NMF Algorithms with alternating least squares update using both regularization parameters as shown in Equation 6.

on one hand the OH component contains energy across the entire spectrum and on the other hand the elemental components are sparse in nature, it is preferable to minimize the energy in the hidden components as well as the reconstruction error. In classical NMF without minimization constrains the most prominent component (in this case OH) ends up being repeated across most of the extracted components. This is because classical NMF only tries to minimized reconstruction error and distributes the dominate component over all extracted components to satisfy a minimal reconstruction error. When minimization constraints are imposed the most prominent component is forced to a fewer number of extracted components since the cost function is now trying to minimize both reconstruction error and the energy in the hidden components. The constant α_1 is user specified and controls the relative weight of this constraint in the overall optimization problem. The presence of this second term creates a tradeoff between the smoothness of **S** components with the reconstruction error. The optimization algorithm will tend to minimize first term (Euclidean distance) of C_1 while minimizing the 2nd term for a given α_1. Therefore, while running the optimization at each iteration, the optimization algorithm tends to scale down **S** while scaling up **a** in return, so that the product of these two terms always stays the same. In the existing literature, the way this problem is handled is by rescaling **S** and **a** after each iteration [9]. This means, after every iteration, each column of **a** is normalized to unit length followed by an update of **S**.

Cost function C_2 has similar properties to C_1 with the added constraint that the energy of **a** is to be minimized as well. This results in enforcing a smoothness constraint on both **S** and **a** while potentially increasing the reconstruction error. For this case, since both terms are included in the optimization function, no rescaling is necessary. The pseudo code of NMF that corresponds to C_2 cost function is given in Figure 3. Here the update rule is obtained from the least square solution of the derivative of the objective function with respect to **S** or **a**. This method is known as "NMF with least square update" and further details of this approach can be obtained in the some of the very recent literatures [3, 13]. In the current scope of this study, we intend to demonstrate the applicability of the above mentioned blind source algorithms to analyze high dimensional spectral time series $Y(t,\lambda)$ in order to detect the presence of element burns in the wavelength-time plane as an indicative of the degradation of the system's heath. In this context it is worth mentioning here that the choice of the second variants of NMF (as given in Figure 3) is very instinctive as we are searching for the sparse failure profiles corresponding to each element from the severity list.

INDEPENDENT COMPONENTS ANALYSIS

Independent Components Analysis (ICA) [7] has received wide-spread attention as a new method of signal

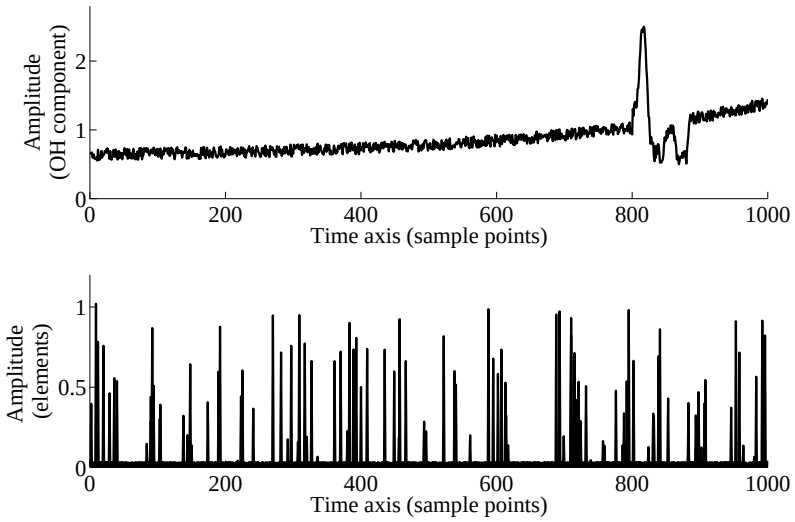


FIGURE 4. The figure shows the true failure profiles corresponding to the OH component and the element burns, plotted across time.

decomposition. It assumes that the signal matrix Y is a superposition of components that are statistically independent and non-Gaussian. Throughout this research, the FASTICA Matlab Toolbox has been used for analysis purpose and for detailed literatures or source code we ask the readers to refer to the following resource (http://www.cis.hut.fi/projects/ica/fastica/index.shtml).

TABLE 2. This table shows the Area under the ROC curve for SDA, PCA, and NMF. Here α_1 & α_2 are the regularization parameters. Note that the SDA algorithm and PCA algorithms have very similar performance.

Algorithms	Area Under ROC		Time Complexity
	mean	std	(Seconds)
SDA	0.93	0.009	11
PCA	0.94	0.000	11
ICA	0.75	0.012	25
NMF (α_1,α_2)	0.86	0.106	1867
NMF (prior)	0.83	0.031	3577
SDA (prior)	0.93	0.000	7

4. RESULTS

Table 2 represents a comparative study on the performance of all the four different algorithms on a spectral data set (Y) which has a dimension of 1061 and 1000 instances. The true temporal profile of the hydroxide (OH) and element burns corresponding to test set (Y) has been shown in Figure 4. Any non-zero activity in the temporal elemental profile is considered a true failure. As mentioned earlier, the outcome of any of these decomposition algorithms is a set of basis vectors and their corresponding failure profiles. In this study, we have deliberately extracted 12 hidden components because we assume that the first 12 basis vectors will contain most of the spectrum information regarding the OH burn, 10 element burns and

background scatter. The temporal and spectral components extracted using SDA and NMF can be seen in Figures 5 and 6. For analysis purpose, we first reconstructed a data matrix $\hat{Y}_{true}$ with only the true element burns and their time profiles. This new data matrix serves as a ground truth representing the varying energy profile corresponding to the element burns over time. Similar data matrix $\hat{Y}_{algo}$ has been constructed using the profiles (spectrum and time) of the decomposed element burns extracted from each individual algorithm. Thereafter a detection threshold has been imposed on $\hat{Y}_{true}$ and $\hat{Y}_{algo}$ to calculate the "area under the ROC curve", a metric that has been used to evaluate the performance of the algorithm in this study. For each algorithm, all the readings corresponding to 50 runs have been recorded. The numbers shown in Table 2 represent the mean and the standard deviation calculated over those 50 runs. The right most column of the table represents the mean time complexity of each algorithm.

From Table 2 it can be seen that both SDA and PCA exhibit similar performance and emerge as winners. We have also observed that standard NMF was unable to separate the OH burns from the element burns and this is understandable as standard NMF would always try to minimize the reconstruction error and it will try to achieve this by distributing the energy of the most dominant feature (in this case the OH burn) over all the 12 extracted components. The NMF algorithm with sparsity factors as proposed by Hoyer [8] was unable to provide with some meaningful solution in this case, as the data matrix (Y) is composed of both non-sparse OH profile and sparse element profiles in wavelength-time domain. However NMF with regularization parameters as found in Equation 6 was able to present a much better performance compared to standard NMF with/without sparsity and ICA while detecting the element burns (refer Figures 6).

In a separate study, we have incorporated domain knowledge in some of the algorithms like SDA and NMF. This was done by initializing the first 10 components of $\mathbf{S}$ with digitalized signal (of 1-s and 0-s) having peaks at the primary and secondary wavelengths corresponding to all 10 element while the rest 2 were initialized randomly. The result showed no particular improvement in the performance of SDA with additional domain knowledge. However there was a noticeable improvement in the time complexity. While standard NMF (Figure 2) did not work in the first place but with the domain knowledge included, the same algorithm proved to be successful in detecting the element burns effectively.

5. Conclusions

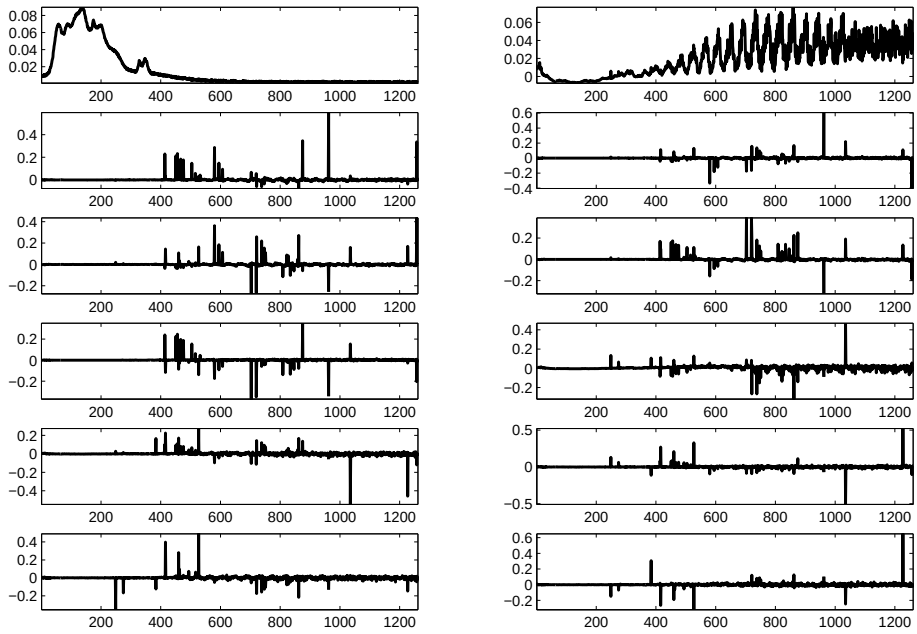
This paper has reviewed some of the most recent and popular blind source separation techniques to generate low dimensional signals, which provide the best description of the hidden features associated with the system states. In this paper we have discussed the use of standard algorithms like SDA, PCA, ICA and NMF to extract hidden features as a necessary step towards anomaly detection on high dimensional data sets and finally provided a comparative study of the performances of these methods under different detection criteria. We have also described a spectral emulator that provides a good approximation of some of the events arising from the plume of the space shuttle main engine and this also serves as a good source of high dimensional data sets. The above mentioned algorithms have demonstrated the ability to detect the presence of element burns and separate them from OH profiles. Furthermore the detection method applied was based on a fixed threshold, which leaves room for improvement in future work where more advanced machine learning algorithms that are not simply amplitude based can do much better at detecting similar types of anomalies.

6. Acknowledgements

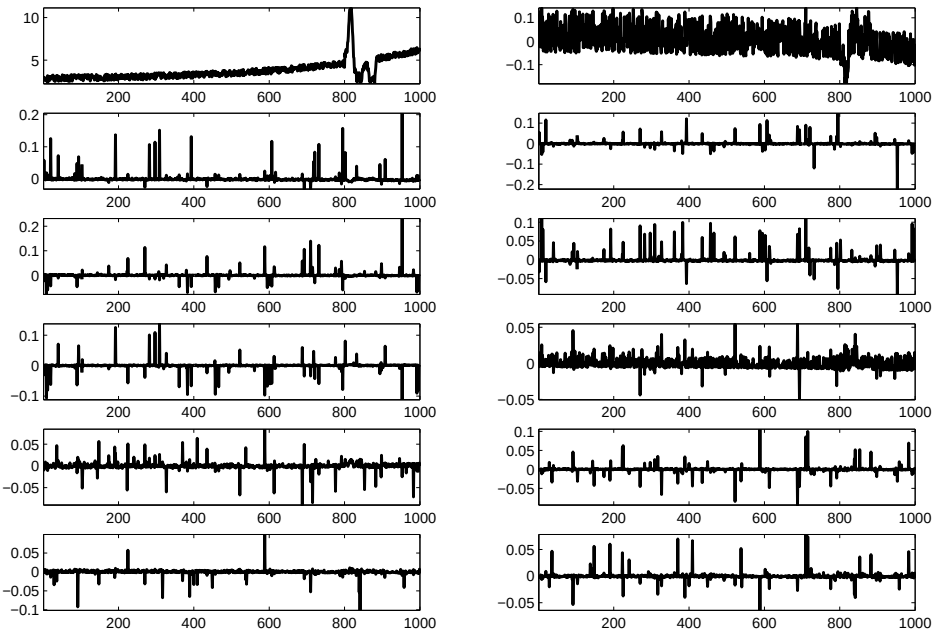
This work was supported through funding from the NASA Aeronautics Research Mission Directorate, Aviation Safety Program, Integrated Vehicle Health Management project. The authors thank Drs. Kanishka Bhaduri and Nikunj Oza for valuable discussions and suggestions.

References

1. Daniel A. Benzing and Kevin W. Whitaker, *Approach to space shuttle main engine health monitoring using plume spectra*, Journal of Spacecraft and Rockets **35** (1998), no. 6, 830–836.
2. Kevin W. Whitaker Daniel A. Benzing and Randall C. Hopkins, *Experimental verification of neural network-based ssme anomaly detection*, American Institute of Aeronautics and Astronautics, Inc. (1997).
3. P. Pauca F. Shahnaz, M. Berry and R. Plemmons, *Document clustering using non-negative matrix factorization*, Information Processing and Management **42** (2006), 373–386.
4. David B. Van Dyke Gopal D. Tejwani and Felix E. Bircher, *Approach to ssme health monitoring iii. exhaust plume emission spectroscopy: Recent results and detailed analysis*, AIAA/SAE/ASME/ASEE 29th Joint Propulsion Conference and Exhibit (1993).
5. David B. Van Dyke Gopal D. Tejwani and Felix E. Bircher, *Ssme exhaust plume emission spectroscopy at ssc: Recent analytical developments and test results*, 31st AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit (1995).
6. Felix E. Bircher Gopal D. Tejwani, David B. Van Dyke and Donald G. Gardner, *Emission spectra of selected ssme elements and materials*, NASA-RP-1286 (1992).
7. T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: Data mining, inference, and prediction*, Springer, 2001.
8. P. O. Hoyer, *Non-negative sparse coding*, 2002, in Proc. IEEE Workshop on Neural Networks for Signal Processing.
9. P. O. Hoyer, *Non-negative matrix factorization with sparseness constraints*, Machine Learning Research **5** (2004), 1457–1469. (Source code can be downloaded from http://www.cs.helsinki.fi/u/phoyer/software.html)
10. I. T. Jolliffe, *Principal component analysis*, Springer, 2002.
11. K. H. Knuth, *A bayesian approach to source separation*, Proceedings of the First International Workshop on Independent Compenents Analysis and Source Separation (1999), 283–288.
12. Daniel D. Lee and H. Sebastian Seung, *Algorithms for non-negative matrix factorization*, NIPS, 2000, pp. 556–562.
13. A. Langville P. Pauca M. Berry, M. Browne and R. Plemmons, *Algorithms and applications for approximate nonnegative matrix factorization*, Computational Statistics and Data Analysis **52** (2007), no. 1, 155–173.
14. C. R. Rao, *Linear statistical inference and its applications*, John Wiley and Sons, 1965.
15. H. Park S. Choi, A. Cichocki and S. Y. Lee, *Blind source separation and independent component analysis: A review*, Neural Information Processing - Letters and Reviews **6** (2005), no. 1, 1–57.
16. A. N. Srivastava and W. Buntine, *Data analysis of components in the optical plume of the space shuttle main engine*, Proceedings of the AIAA (1995).
17. G. D. Tejwani, D. B. van Dyke, F. E. Bircher, and D. J. Chenevert, *Emission spectra of selected ssme elements and materials*, (1992), Technical Report: NASA-RP-1286, Document ID: 0133990.

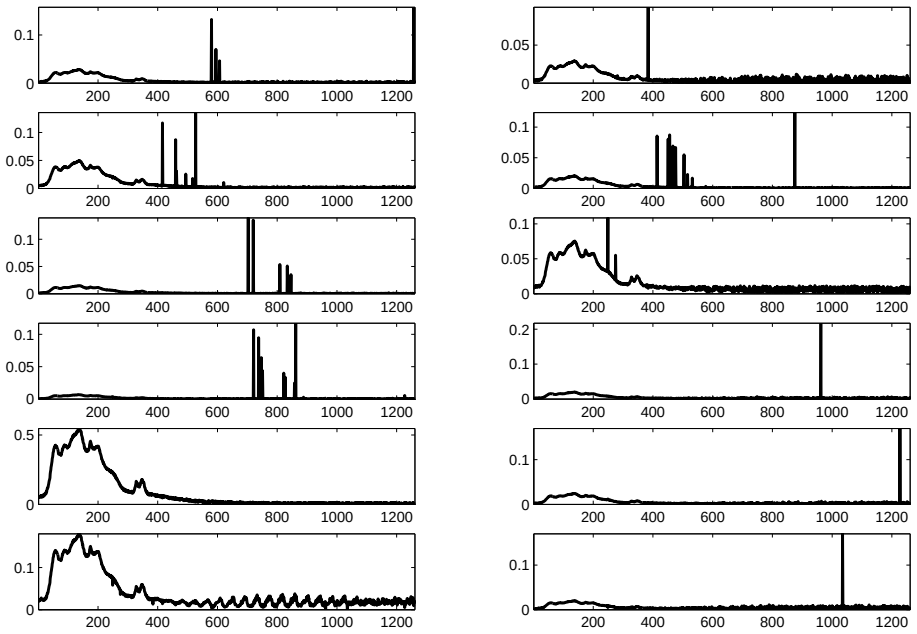


(a) Spectrum profile

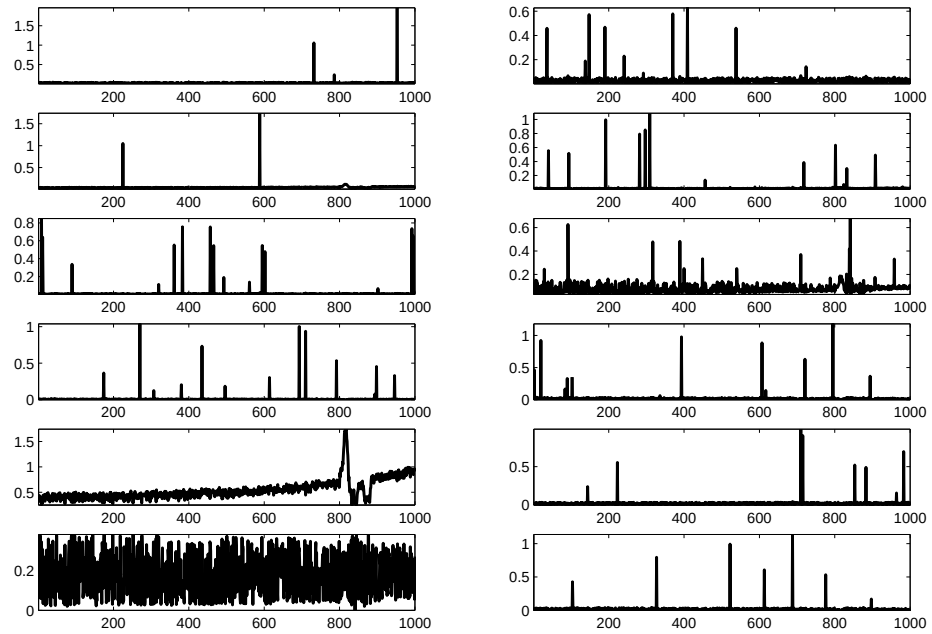


(b) Time profile

FIGURE 5. The figure shows the outcome of the SDA algorithm without any prior knowledge. Figure 5(a) represents the extracted basis vectors while Figure 5(b) represents the time profile of each basis vectors, arranged in the same sequence.



(a) Spectrum profile: A total of 12 extracted basis vectors



(b) Time profile of individual basis vector

FIGURE 6. The figure shows the outcome of the NMF algorithm with regularization terms as expressed in Equation 6. In this analysis, the regularization parameters α_1 and α_2 are set to 0.01 and 0.001 respectively.